Реферат прочитан на Научниот собир
"IX денови на Благоја Корубин",
секција "Информатиката и лингвистиката",
Институт за македонски јазик, Скопје 30-31.03.2005

ЗВОНКО НИКОДИНОВСКИ

ЗА ЕДЕН СЕОПФАТЕН ЕЛЕКТРОНСКИ РЕЧНИК НА МАКЕДОНСКИОТ ЈАЗИК

Можеби изгледа малку нереално да се зборува за сеопфатен електронски речник на македонскиот јазик во услови кога не постои ни целосен толковен речник на македонскиот јазик. Сепак, сметаме дека е потребно и корисно да се дискутира за потребите, за можностите и за идеите преку кои може да се дојде до остварување на посакуваната цел.

Нашето излагање ќе го сосредоточиме на две теми: 1. Претпоставки за реализација на проектот и 2. Проектот "Сеопфатен електронски речник на македонскиот јазик".

I. Претпоставки за реализација на проектот

Во врска со претпоставките за реализација на проектот можеме само да кажеме дека, според наши сознанија, во моментот не постојат тимови кои можат заеднички да работат на еден ваков проект. Проектот за еден сеопфатен електронски речник на македонскиот јазик подзразбира здружување на силите на неколку видови истражувачи : лингвисти, информатичари и документалисти. Информатичарите кои работат на Институтот за информатика, на Електротехничкиот факултет или пак во разни фирми од областа на информатичките услуги се препуштени сами на себе и на потребите на пазарот. А пазарот е таков што бара апликации кои имаат директна примена во фирмите (водење на сметководство, продажба и магацинско работење, бази на податоци од разен вид и др.) и немаат директна врска со потребите на јазичните индустрии.

Нашите сограѓани досега се задоволуваа со производите на светските компании на големите странски јазици (англиски, француски, германски и др.) или пак на јазиците на нашите соседи (српски, хрватски и др.). Познатите информатички фирми веќе почнаа, иако во минимална степен, да го застапуваат и македонскиот јазик во своите производи (на пр. можноста за страница на македонски јазик кај Google, или пак присутноста на македонскиот јазик во програмата за оптичко препознавање на карактери Fine Reader и др.).

Во последните неколку години некои наши софтверски компании (Семос, Логин системи) почнуваат да мислат на домашниот корисник нудејќи му производи и на македонски јазик (програма за делење на слогови и коректор за македонски јазик, конвертор за фонтови и др.). Треба да се спомене исто така и производството на цд-ромови од разен тип (енциклопедии, текстуални и др.) кои сосема малку ја покриваат таа огромна празнина во македонскиот информатички простор.

На полето на македонскиот јазик или пак на македонско-инојазичните електронски помагала, ситуацијата е поразителна. Освен еден англиско-маке-

донски речник кој нема речиси никаква употреба кај луѓето кои работат секојдневно со компјутери, речиси не постои ниедно позначајно помагало во областа на јазичните електронски помагала.

Лингвистите, од своја страна, работат сами и сонуваат своите дела да ги видат во пишана форма, на хартија. Ниту лингвистите, ниту издавачките куќи немаат пријавено интерес за електронско издавање. А бројот на млади луѓе кои од ден на ден се користат со компјутерите како со основно средство за работа е сè поголем и денеска речиси нема млад човек кој на некој начин (преку образованието или преку интернетските кафулиња) не бил во ситуација да го употреби компјутерот. За жал, сè уште е голем бројот на оние кои сметачите ги користат само како уреди за играње. Одоворноста за тоа ја сносат наставниците кои не ги упатуваат учениците во тој домен, потоа родителите кои во голем дел не се во состојба да ги водат децата во нивното користење на компјутерите и на крај информатичарите кои не успеале да им понудат на младите интересни прозводи на мајчин јазик.

Лингвистите треба да почнат да размислуваат за тоа дека преку компјутерите можат да се создадат дела од разен вид кои на денешните генерации и посебно на младите им се потребни за да се образуваат и да бидат во состојба да се натпреваруваат со своите врсници од другите земји во светот затоа што денешната комуникација е пред сè електронска комуникација. А младите тоа добро го знаат имајќи предвид дека речиси нема млад човек кој не користел уште од тинејџерските години мобилен телефон.

Пред да почнеме со набројување на задачите, би сакале да го искажеме нашето уверување дека во денешно време не може да постои напредок на науката, техниката и воопшто на знаењето ако тие не се достапни на интернет. Токму затоа сметаме дека сè што може да се направи на полето на јазичните индустрии треба да биде бесплатно и достапно за сите. Нашата нација не е толку многубројна и моќна за да може да се наметнува со други производи на светскиот пазар на знаење, технологии и култура. Единствено преку нашите културни производи во поширока смисла кои ќе им бидат достапни на сите луѓе во светот ќе можеме донекаде да се претставиме и да се мериме со другите. Тоа нѝ наложува на сите нас, пред сè на владата и на сите институции во Македонија, безрезервно да се отвориме и да го публикуваме на интернет севкупното богатсво од наследството на Македонија и на македонскиот народ (што би значело на пр. објавувањето на интернет на старите македонски ракописи за афирмација на македонското културно богатство може само да се замисли).

Еве што можеме да предложиме да се направи во Македонија на полето на информациско-коминикациските технологии за потребите на македонската јазична индустрија :

- Да се воведе предметот Информатичка лингвистика на Филолошките факултети во земјата или Лингвистичка информатика на Институтот за информатика.

- Да се формираат Одделенија за информатичка лингвистичка во Институтот за македонски јазик и во Македонската академија на науките и уметностите.
- Да се формира една студиска група за информатичка лингвистика (или за лингвистичка информатика) која ќе биде предложена истовремено на Филолошкиот факултет и на Институтот за информатика и на која ќе се запишуваат студенти од Филолошкиот факултет и од Институтот за информатика.
- Да се работат заеднички научни проекти во кои ќе учествуваат лингвисти и информатичари.
- Да се дигитализира фондот на книги и публикации на македонски јазик што го поседуваат институциите во Македонија и тоа : НУБ " Св. Климент Охридски", Филолошкиот факултет "Блаже Конески", МАНУ, Институтот за фолклор, Институтот за македонска литература, Институтот за национална историја, Архивот на Македонија и сите други и библиотеки и архиви во Македонија, Фондот на Македонската радио-телевизија и на другите медиски куќи во Македонија, фондовите на дневните весници и списанија и др. Дигитализацијата да се одвива истовремено на македонските пишани и говорни материјали.
- Сиот дигитализиран материјал да биде достапен на интернет преку веб страници на соодветните институции. Секоја институција да изработи и библиорафија на дигитализираните материјали што ги поседува и да ги даде истите на користење или симнување преку мрежата интернет.
- Универзитетите во Македонија како и НУБ "Св. Климент Охридски" да почнат да ги објавуваат на интернет дипломските, магистерските и докторските трудови и да имаат обврска во иднина да ги објават сите трудови одбранети на соодветната институција.
- На Филолошкиот факултет "Блаже Конески" и во Институтот за македонски јазик да се формираат електронски корпуси кои ќе бидат постојано ажурирани.
- Истовремено со електронските корпуси да се направат и програми за пребарување, анализа и користење на тие корпуси.
- Катедрата за македонски јазик да почне со работа врз електронски методи за изучување на македонскиот јазик како странски јазик.
- Катедрите за странски јазици и за јазиците на етничките заедници во Македонија, од една страна, и соодветните стручни институции (факултети, институти и др.) да формираат преведувачки меморији за соодветните парови јазици во однос на македонскиот. Во работата врз формирање на тие меморији треба да учествуваат лингвисти и специјалисти од дадените области (пр. инжинери, лекари, архитекти, географи и др.). Врз база на тие меморији да почнат да се формираат истовремено терминолошки

речници за дадените области. Сите материјали, од меморii до речници, да бидат достапни на интернет.

- Катедрата за македонски јазик, Институтот за македонски јазик и идното Одделение за информатичка лингвистика при МАНУ да почнат со информатички описи на структурата на македонскиот јазик што ќе доведе до разни електронски речници.
- Да почнат да се локализираат одредени компјутерски апликации на македонски јазик. Министерството за образование и наука да финансира изготвување на разни програми од обласа на јазичните индустрии : програми за искористување на преведувачките меморii и на корпусите, јазични коректори, програми за препознавање и синтеза на говор (за македонскиот јазик), програми за оптичко препознавање на карактери, програми за правење на разни видови речници, програми за автоматско преведување за македонски јазик.

II. Проектот "Сеопфатен електронски речник на македонскиот јазик".

Во овој дел од излагањето сакаме да го изнесеме нашето видување за тоа што треба да содржи Сеопфатниот електронски речник на македонскиот јазик. Се разбира дека при тоа ќе се служиме со постојните електронски речници на одредени јазици (француски, англиски, италијански, шпански, германски, романски, српски, словенечки, бугарски) од кои ќе го земеме она што го сметаме за корисно на што ќе му додадеме и нешто што може да го направи проектираниот речник уште поинтересн и покорисен.

Која ќе биде целта на Сеопфатниот електронски речник на македонскиот јазик ? Сеопфатниот електронски речник треба да биде речник на стандардниот македонски јазик кој треба да го содржи сето богатство за македонскиот јазик со кое располагаат македонските лингвисти.

Главниот дел на речникот се состои од речнички статии кои за заглавни ќе ги имаат основните форми на македонскиот јазик : лексемите (прости и сложени) и формативните елементи (префикси, суфикси, конфикси). Освен општата лексика, речникот ги содржи и карактеристичните македонски лични имиња, презимиња, прекари и топоними (со опис на нивните составни делови).

Фондот на општата лексика го опфаќа стандардниот македонски јазик во сите свои разновидности (стандардниот јазик, разговорниот јазик и стручните терминологии). Дијалектните разновидности се појавуваат во статиите посветени на топонимите или на одредниците *дијалект, говор, наречје*.

Речникот претставува збир на бази на податоци од која со помош на разни пребарувувања се добиваат резултати во вид на формулари. Речникот ги содржи следните бази : 1. база на леми 2. база на флексивни форми 3. база на хомоними 4. база на синоними 5. база на антоними 6. база на фраземи 7. база на пословици 8. база на цитати (со афоризми) 9. база имиња и презимиња 10. база на топоними

Базата од леми е формирана од записи од кои секој ја содржи лемата (заглавниот збор) и полиња кои ја опишуваат лемата. Базата ги содржи следните полиња : 1. лема 2. изговор на лемата 3. фонетска транскрипција со акцент 4. категорија на лемата 5. правописна забелешка 6. латинична форма на лемата 7. етимологија на лемата со датација 8. значење I на лемата 9. област на значење I 10. пример на значење I 11. значење II на лемата со ознака фигуративно значење и видот на фигурата 12. област на значење II 13. пример на значење II.